Neuberger, A. (2012). Realized Skewness. The Review of Financial Studies, 25(11), pp. 3423-3455. doi: 10.1093/rfs/hhs101

REALIZED SKEWNESS

Anthony Neuberger

Warwick Business School
Warwick University
Coventry
CV4 7AL
United Kingdom

(44) 24 7652 2995

anthony.neuberger@wbs.ac.uk

The author is grateful to comments from an anonymous referee, the Associate Editor, Pietro Veronesi, and from Carol Alexander, Mark Britten-Jones, Peter Carr, Pat Forde, David Hobson, Roman Kozhan, Eberhard Mayerhofer, Alessandro Palandri, Johannes Rauch, Thomas Ruf, Paul Schneider, Neil Shephard, Hao Zhou and seminar participants at the UBC Summer Conference, the Frankfurt MathsFinance Conference, the Oxford-Man Institute, Baruch College, the Universities of Maryland, Reading, Piraeus, and Warwick, Morgan Stanley, and the Federal Reserve Board.

The third moment of returns is important for asset pricing, but it is hard to measure precisely, particularly at long horizons. This paper proposes a definition of the realized third moment that is computed from high frequency returns. It provides an unbiased estimate of the true third moment of long horizon returns, doing for the third moment what realized variance does for the second moment. The methodology is used to demonstrate that the skewness of equity index returns, far from diminishing with horizon, actually increases with horizons up to a year, and its magnitude is economically important.

The third moment of returns is important for asset pricing. But the third moment, particularly of long horizon returns, is hard to measure precisely. This paper proposes a definition of the realized third moment that is computed from high frequency returns and from option returns. It provides an unbiased estimate of the true third moment of long horizon returns. The novel methodology is used to demonstrate that the skewness of equity index returns, far from diminishing with horizon, actually increases with horizons up to a year, and its magnitude is economically important.

While standard approaches to asset pricing concentrate largely on the first and second moment of returns, there is mounting evidence that higher moments are also important. The literature going back to Kraus and Litzenberger (1976), and including more recently Harvey and Siddique (2000), Ang, Hodrick, Xing and Zhang (2006), Ang, Cheng and Xing (2006) and Xing, Zhang and Zhao (2010), suggests that the asymmetry of the returns distribution both for individual stocks and for the market as a whole is important for asset pricing and investment management. Skewness is central to the debate on the role of large rare disasters in explaining the equity risk premium (Rietz (1988), Longstaff and Piazzesi (2004), Barro (2009), and Backus, Chernov and Martin (2011)). Carr and Wu (2007) document the time varying implied skew in foreign exchange markets, and Brunnermeier, Nagel and Pedersen (2008) relate the forward premium puzzle to the skewed distribution of currency returns.

Investigating the impact of skewness on asset pricing is hard because the technology for measuring the skewness of returns at long horizons is not well developed. Estimates of the higher moments of distributions are plagued by noise and heavily influenced by outliers (Kim and White, 2004, and also section 1 of this paper); with long horizon returns the problem is compounded by the small number of non-overlapping observations. The empirical literature has side-stepped the problem by focusing on skewness and co-skewness at monthly or higher frequencies. But skewness, unlike variance, does not scale nicely with horizon, and it is not clear what the relationship is between skewness at short and long horizons, or why asset prices in an economy with well-capitalized long term investors should be heavily influenced by the characteristics of short horizon returns.

Skewness in long horizon returns comes from two sources: the skewness of short horizon returns and the correlation between returns and volatility innovations (the leverage effect). In this paper, I show how to use high frequency[1] returns to compute the third moment of long horizon returns. The computation uses option return data as well as returns on the underlying asset. The option returns help capture the leverage effect. The existence of strong negative correlation in the equity market has been much discussed since being documented by Black (1976) and Christie (1982), and turns out to be much more important than the skew in high frequency returns in delivering skew in index returns at long horizons.

Because of the close analogy with the use of high frequency returns to compute realized variance, I call the computed quantity the realized third moment. The realized third moment is an unbiased estimate of the true third moment. The lack of bias is not based on any model; it only requires that prices are martingale. It is robust to price jumps and discrete sampling. I show that this property of freedom from bias is sufficient to define the realized third moment uniquely. The realized third moment can then be normalized by dividing by the variance to the power of 3/2, to give the realized skewness coefficient.

If there are risk premia, the martingale assumption is violated and the realized third moment may be biased. I quantify the extent of this bias both theoretically and in simulations.

As a by-product of the derivation of the realized third moment, a new definition of realized variance is also proposed; unlike the standard definition, it provides an unbiased estimate of true variance, even in the presence of jumps. The parallel with variance goes further. Just as there is a model-free strategy to replicate a variance swap, a swap where the floating leg is the realized variance, so there is a model-free strategy to replicate a skew swap, one where the floating leg is the normalized realized third moment. Carr and Wu (2009) use variance swaps to explore variance risk premia; Kozhan, Neuberger and Schneider (2011) use skew swaps to explore the existence and behavior of risk premia associated with skew.

I show, using daily data, that the skew in the equity index market, far from declining with horizon as would occur under an iid process, is actually higher at one year than it is at one month. Furthermore the increased precision of measurement makes it possible to document that the skew in index returns is time varying, and that the degree of variation is economically significant.

These findings are relevant to the debate on the role of large disasters in explaining the equity premium. In reviewing the evidence, Backus, Chernov and Martin (2011, p1970, “BCM”) note that “in virtually all of this research [starting with Rietz (1988), followed by Longstaff and Piazzesi (2004), Barro (2009) and others], the distribution [of log returns] is modeled by combining a normal component with a jump component. The jump component, in this context, is simply a mathematical device that produces nonnormal distributions”.

The parameters chosen by BCM for the process for the equity index (under the physical measure) imply that the skewness of daily returns is -0.6. This is consistent with US data. But, coupled with the iid assumption, it implies a skewness of annual returns of -0.04

(see BCM, Table 3), in contrast with a figure of around -1 suggested by the analysis presented here. If the horizon of the representative agent is of the order of years rather than days, the calibration of the model to daily returns may understate the role that rare disasters play in asset pricing. High frequency models used to investigate large disasters need to capture the dependency between returns and variance over long horizons that is observed in the data.

Previous work on trading the skew includes Schoutens (2005) and Schoutens, Simons, and Tistaert (2005) who describe swap contracts that pay the sum of cubed daily returns. The swaps capture the third moment of high frequency returns, but do not take account of the leverage effect that heavily affects the third moment of returns over the life of the swap. Bakshi, Kapadia and Madan ("BKM", 2003) show how the implied moments and skewness coefficient of the risk-neutral probability density can be recovered from option prices. The implied density in general differs substantially from the density under the physical measure (see Carr and Wu, 2009). There is no obvious way of relating the implied BKM skew at any horizon to the actual behavior of high frequency returns up to that horizon.

The paper is organized as follows: section 1 presents empirical evidence on the skew in equity index returns to motivate the rest of the analysis in. Section 2 derives realized skewness for price changes. Section 3 extends the theory to apply to returns. The main theoretical results – the characterization of all realized moment estimators (Proposition 2), the definition of the realized third moment (Proposition 6), and the bias when prices are not martingale (Proposition 7) - are contained in this section. The precision and bias of realized skewness are investigated through simulation in section 4, and the methodology is taken direct to the data to investigate the term structure of skewness of index returns in section 5. The final section concludes.

1. Skewness of Equity Index Returns

Daily returns on the S&P500 are on average negatively skewed. The skewness coefficient of excess log returns[2] over the period July 1963 to June 2011 is -0.83. The result appears to be significant; the *p*-value against the null that the skew is positive is 4.3%.

But even with a very long run of data, the precision of the estimate is low, and the point estimate is heavily influenced by outliers. Specifically, the conclusion that daily returns are negatively skewed relies heavily on one observation – 19 October 1987. The top panel of Figure 1 shows the skewness coefficient and the 5/95% confidence intervals computed from rolling ten-year windows. The impact of the 1987 Crash is obvious. For most of the windows (including periods affected by the Crash) the hypothesis that the skew is positive cannot be rejected at conventional significance levels. The weakness of the evidence for negative skewness of daily returns has been noted by other authors including Kim and White (2004).

When the same technology is applied to monthly returns over the same period, the conclusions are somewhat different. The point estimate for the period as a whole is -0.74; the *p*-value is 0.1% suggesting strong evidence for negative skewness. The lower panel of Figure 1 shows the rolling ten-year estimates for monthly returns. The 1987 Crash does not stand out. While the point estimates from the rolling estimates are generally negative, the data only rejects the null of zero or positive skewness in the early and later parts of the period; in most of the middle period, the null cannot be rejected.

One other striking feature is the high level of skewness in monthly relative to daily returns. If stock returns were iid, the moments of n-day returns would be proportional to n and the skewness coefficient would be proportional to $1/\sqrt{n}$. There is no evidence of such a decline in skewness with horizon in the data. To explore the relation between skewness coefficients of daily and monthly returns more fully, I bootstrap the daily returns to generate a synthetic time series, and compute the population skewness of both daily and monthly returns. The results are plotted in the top panel of Figure 2. The lower panel does the same thing using monthly and annual rather than daily and monthly returns.

The hypothesis that returns are iid and the skew at the monthly horizon derives from the skew at the daily horizon is firmly rejected by the data (*p*-value of 0.02%), and the hypothesis that annual skew is generated by monthly skew is also rejected (*p*-value of 4.76%). This is clear evidence that the distribution of low frequency returns is heavily influenced by serial dependence in high frequency returns. If high frequency returns are to be used to improve the estimate of the skewness of low frequency returns, it must be done in a way that reflects the serial dependencies that are manifest in the data.

2. Arithmetic Contracts

This section introduces the Aggregation Property – the property that ensures that the quantity measured using high frequency returns is an unbiased estimate of its low frequency counterpart. The goal is to get a good measure of the third moment of returns, and that will be achieved in section 3. This section looks at price changes rather than returns because the mathematics are simpler, and the underlying logic more transparent.

2.1 Notation and Terminology

S_t $\left(t \in [0,T]\right)$ is a positive adapted variable defined on a standard filtered probability space $\left(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\in[0,T]}, \mathbb{P}\right)$. $\mathbb{E}_t[.]$ denotes $\mathbb{E}\left[.|\mathfrak{F}_t\right]$. The distribution of S_T is assumed such that expectations of S_T and functions of S_T such as S_T^2, ln(S_T) and S_T ln(S_T) exist. $\mathbf{T} = \{t_0 = 0 < t_1 < ... < t_N = T\}$ is a partition of [0,T]. $\|\mathbf{T}\| = \max_i \{t_i - t_{i-1}\}$ is the mesh of **T**. For a process x, x_i is shorthand for x_{t_i}. δx_i denotes $x_i - x_{i-1}$. For a function g(.), $\sum^{\mathbf{T}} g(\delta x)$ is shorthand for $\sum_{i=1}^{N} g(\delta x_i)$

I will sometimes refer to the period [0,T] as a month, and the length of the sub-period as a day, but obviously nothing hangs on this. In the present section, references to variance and skewness relate to the distribution of price changes and not of returns. To avoid irrelevant complications with interest rates and dividends, I work throughout with forward prices, so all trades, whenever entered into, are for settlement at time T.

2.2 The Aggregation Property

For any martingale S

$$\mathbb{E}_0\left[\left(\sum^{\mathbf{T}} \delta S\right)^2\right] = \mathbb{E}_0\left[\sum^{\mathbf{T}} (\delta S)^2\right]. \tag{1}$$

There are several interpretations of this equation. The left hand side is $\mathbb{E}_0\left[(S_T - S_0)^2\right]$. With S a martingale, this is equal to $\mathbb{E}_0\left[\left(S_T - S_0 - \mathbb{E}_0[S_T - S_0]\right)^2\right]$, the variance of the monthly price change S_T-S_0 . If the probability measure used is the physical measure, this is the *true variance* of *S*. Equation (1) then says that if prices are martingales, the sum of squared daily price changes (the *realized variance*) is an unbiased estimator of the true variance. If the measure is a pricing measure, it says that the fair price of a one-month variance swap computed daily (a swap that pays the realized daily variance over a month) is the same as the price of a contingent claim that pays $\left(S_T - S_0\right)^2$. Indeed, since the relationship holds under any pricing measure (since the process is martingale under any pricing measure), it also implies that a variance swap can be perfectly replicated if the contingent claim exists (or can be synthesized from other contingent claims), and the underlying asset is traded. It is reasonable to call the time 0 price of the claim the *implied variance*.

The relation between the true variance of the monthly price change, the realized variance of daily changes over the month and the implied variance of one-month options at the beginning of the month holds exactly, whatever the price process and whatever the length and number of sub-periods, provided only that *S* is a martingale. It depends on the interchangeability of the summation and the square function under the expectations operator.

In order to generalize the notion of true, realized and implied characteristics, some more definitions are needed. If *g* is a real-valued function and *X* is an adapted (scalar or vector) process, then (*g*; *X*) has the *Aggregation Property* if, for any times $0 \le s \le t \le u \le T$,

$$\mathbb{E}_s\left[g\left(X_u - X_s\right)\right] = \mathbb{E}_s\left[g\left(X_u - X_t\right)\right] + \mathbb{E}_s\left[g\left(X_t - X_s\right)\right]. \tag{2}$$

Applying the Law of Iterated Expectations, if (*g*; *X*) has the Aggregation Property then $\mathbb{E}_0\left[g\left(X_T - X_0\right)\right] = \mathbb{E}_0\left[\sum^{\mathbf{T}} g\left(\delta X\right)\right]$ for any partition **T**. $\sum^{\mathbf{T}} g\left(\delta X\right)$ is the *realized* characteristic; and $\mathbb{E}_0\left[g\left(X_T - X_0\right)\right]$ is the *implied* characteristic if the measure is a pricing measure, or the *true* characteristic if the measure is the physical measure.

(g; S) has the Aggregation Property for any martingale S when $g(x) = x^2$. Proposition 1 below shows that no other interesting functions have the Aggregation Property. But there is no reason to require that $X = S$. Let $X(S)$ be a vector-valued process $\left\{S_t, V_t\left(S_T\right) : t \in \left[0, T\right]\right\}$ where $V_t\left(S_T\right) = Var_t\left[S_T\right]$, the variance of S_T conditional on information at t. **G** is the set of analytic functions g such that $\left(g; X\left(S\right)\right)$ has the Aggregation Property for all martingales S.

Proposition 1: G *consists of the functions*

$$g\left(\delta S, \delta V\right) = h_0 \delta V + h_1 \delta S + h_2 \left(\delta S\right)^2 + h_3\left(\left(\delta S\right)^3 + 3\delta S \delta V\right). \tag{3}$$

where the $\{h_i\}$ are arbitrary constants.

Proof: see Appendix 1.

G is spanned by four functions: $g_0 = \delta V$ and $g_1 = \delta S$, which are uninteresting, $g_2 = \left(\delta S\right)^2$, which is the variance and is familiar, and $g_3 = \left(\delta S\right)^3 + 3\delta S \delta V$. The true characteristic of g_3 is

$$\begin{aligned}
\mathbb{E}_0\left[g_3\left(S_T - S_0, V_T - V_0\right)\right] &= \mathbb{E}_0\left[g_3\left(S_T - S_0, -V_0\right)\right] \text{ (since } V_T = 0\text{)} \\
&= \mathbb{E}_0\left[\left(S_T - S_0\right)^3 - 3V_0\left(S_T - S_0\right)\right] \text{ (definition of } g\text{)} \\
&= \mathbb{E}_0\left[\left(S_T - S_0\right)^3\right] \text{ (since } S \text{ is a martingale).}
\end{aligned} \tag{4}$$

Hence, the characteristic captured by g_3 is the third moment. The left hand side of equation (4) is the true third moment of the price change over the month. The realized third moment is $\sum^{\mathbf{T}} g_3\left(\delta S,\delta V\right)=\sum^{\mathbf{T}}\left(\left(\delta S\right)^3+3\delta S\delta V\right)$. The Aggregation Property means that the realized third moment equals the true third moment in expectation

$$\mathbb{E}_0\left[\sum^{\mathbf{T}}\left(\left(\delta S\right)^3+3\delta S\delta V\right)\right]=\mathbb{E}_0\left[\left(S_T-S_0\right)^3\right]. \tag{5}$$

Equation (5) is significant in several respects. It shows that skewness in low frequency returns derives only in part from the skewness in high frequency returns. The second source (and indeed the only source when S is a continuously sampled continuous martingale) is the covariation between shocks to the price level and shocks to future variance.

It also shows how high frequency data can be used to provide more efficient estimates of the skewness in price changes over a period. This improvement rests on two assumptions: the discounted price process is a martingale, and the variance of the terminal price is in the observer's information set. Third, if a pricing measure is used, the right hand side is the implied third moment (which can be inferred from the prices of options that mature at time T). The difference between the implied and realized third moments can be used to detect and analyze risk premia associated with skewness.

Fourth, again interpreting the equation under a pricing measure, it suggests how a third moment swap can be designed and replicated. A third moment swap pays the difference between the implied and the realized third moments. To replicate the swap, both S and V have to be tradable as well as observable. To make V tradable, assume the existence of a traded square contract, a security that pays S_T^2. Under the pricing measure, the price of the contract on day i is

$$
\begin{aligned}
P_i &= \mathbb{E}_i\left[S_T^2\right] = S_i^2 + \mathbb{E}_i\left[S_T^2 - S_i^2\right] \\
&= S_i^2 + V_i \text{ since } S_i = \mathbb{E}_i\left[S_T\right].
\end{aligned} \tag{6}
$$

The gain from holding one square contract for one period is

$$
\delta P_{i+1} = \left(\delta S_{i+1}\right)^2 + 2S_i \delta S_{i+1} + \delta V_{i+1}. \tag{7}
$$

Suppose an agent enters into a one month third moment swap on day 0, paying floating and receiving fixed, with the realized third moment being computed from daily prices. She uses the fixed payment to buy a contract that pays the cube of the price change over the month. She hedges by holding -3(S_i-S_0) square contracts and $3\left(\left(S_i - S_0\right)^2 - V_i\right)$ forward contracts over day i. If her initial wealth $W_0 = 0$, her terminal wealth is

$$
\begin{aligned}
W_T &= \left\{\left(S_T - S_0\right)^3\right\} - \left\{\sum^{\mathbf{T}} \left(\delta S\right)^3 + 3\delta S \delta V\right\} \\
&- 3\left\{\sum_{i=0}^{N-1} \left(S_i - S_0\right)\left(\left(\delta S_{i+1}\right)^2 + 2S_i \delta S_{i+1} + \delta V_{i+1}\right)\right\} + 3\left\{\sum_{i=0}^{N-1} \left(\left(S_i - S_0\right)^2 - V_i\right)\delta S_{i+1}\right\}.
\end{aligned} \tag{8}
$$

The right hand side of equation (8) is identically equal to zero, shows that the agent can hedge the swap exactly for any finite partition **T**. So in particular it works perfectly with discrete monitoring and with jumps in the price of the underlying or in its variance. The only requirements are that the market is frictionless, and that the asset and the square contract on it can be traded at any time t that is used for computing the realized third moment.

3. Geometric Contracts

3.1 Generalized Variance

Financial economists are interested in the behavior of returns, not price changes. It is tempting to apply the theory in the previous section directly to the log price, $s_t \equiv \ln S_t$. But the log price is not a martingale either under a pricing measure or, in general, under the physical measure. In looking for a definition of implied and realized variance of returns, I start from the premise that it is important to keep the Aggregation Property. The price for this is a relaxation of the definition of variance.

Let f be an analytic function on the real line with the property that $\lim_{x\to 0} f(x)/x^2 = 1$. Given a process s, define the process $v_t^f(s_T) = \mathbb{E}_t\left[f(s_T - s_t)\right]$. I will call $v_t^f(s_T)$ a *generalized variance* process for *s*.[3]

The variance measures that are widely used by academics and practitioners (squared net returns and squared log returns) conform to the definition of generalized variance. Two other generalized variance measures, v^L and v^E, turn out to be important

$$
\begin{aligned}
v_t^L &= \mathbb{E}_t\left[L(s_T - s_t)\right] \text{ where } L(x) = 2\left(e^x - 1 - x\right), \\
v_t^E &= \mathbb{E}_t\left[E(s_T - s_t)\right] \text{ where } E(x) = 2\left(xe^x - e^x + 1\right).
\end{aligned} \tag{9}
$$

To explain the use of the letters *L* and *E*, rewrite (9) as

$$
\begin{aligned}
v_t^L &= 2\mathbb{E}_t\left[\frac{S_T}{S_t} - 1 - \ln\left(\frac{S_T}{S_t}\right)\right] \text{ so } \mathbb{E}_t\left[\ln S_T\right] = \ln S_t - v_t^L/2; \\
v_t^E &= 2\mathbb{E}_t\left[\frac{S_T}{S_t}\ln\left(\frac{S_T}{S_t}\right) - \frac{S_T}{S_t} + 1\right] \text{ so } \mathbb{E}_t\left[S_T \ln S_T\right] = S_t \ln S_t + S_t v_t^E/2.
\end{aligned}
\tag{10}
$$

In a Black-Scholes world, where the underlying asset has constant volatility σ, the price of a log contract, one that pays $\ln S_T$, is $\ln S_t - \sigma^2\left(T-t\right)/2$, so v_t^L is the implied Black-Scholes variance of the log contract. It is the same as the model-free implied variance (MFIV) of Britten-Jones and Neuberger (2000). Similarly, v_t^E is the implied Black-Scholes variance of the contract that pays $S_T \ln S_T$. I call it entropy because of the functional similarity to entropy as used in thermodynamics and information theory.

The following properties of the log and entropy variance will be useful. From the first line of (10)

$$
\begin{aligned}
&v_t^L = 2s_t - 2\mathbb{E}_t\left[s_T\right] \text{ so} \\
&\mathbb{E}_t\left[\delta v_{t+1}^L - 2\delta s_{t+1}\right] = 0.
\end{aligned}
\tag{11}
$$

Similarly, from the second line

$$
\begin{aligned}
&e^{s_t}\left(v_t^E + 2s_t\right) = 2\mathbb{E}_t\left[s_T e^{s_T}\right] \\
&\text{so } e^{s_t}\mathbb{E}_t\left[e^{\delta s_{t+1}}\left(v_t^E + \delta v_{t+1}^E + 2s_t + 2\delta s_{t+1}\right) - v_t^E - 2s_t\right] = 0 \\
&\text{which implies that } \mathbb{E}_t\left[e^{\delta s_{t+1}}\left(\delta v_{t+1}^E + 2\delta s_{t+1}\right)\right] = 0.
\end{aligned}
\tag{12}
$$

3.2 Aggregation with Returns

Let **G*** be the set of analytic functions g where $\left(g;X\left(S\right)\right)$ has the Aggregation Property for any positive martingale S, where $X(S)$ is the vector process (s, v) with $s=\ln S$ and v is a generalized variance process for s.

Proposition 2: *the set* **G*** *comprises the following functions*

$$\begin{aligned}
&g\left(\delta s,\delta v\right)=h_1\delta s+h_2\left(e^{\delta s}-1\right)+h_3\delta v+h_4\left(\delta v-2\delta s\right)^2+h_5\left(\delta v+2\delta s\right)e^{\delta s}\\
&\text{where } \left\{h_i\right\} \text{ are arbitrary constants, with the following constraints:}\\
&\quad -\ \text{if } h_4\neq 0, v=v^L \text{ and } h_5=0;\\
&\quad -\ \text{if } h_5\neq 0, v=v^E \text{ and } h_4=0;\\
&\quad -\ \text{if } h_4=h_5=0, v \text{ is any generalized variance measure.}
\end{aligned}\qquad (13)$$

Proof: see Appendix 1.

I now examine the properties of three particular members of **G***.

Proposition 3: *the function* $g^M\left(\delta s\right)\equiv e^{\delta s}-1$ *is a measure of expected return and has the Aggregation Property.*

Proof: The Aggregation Property follows immediately from Proposition 2 with $h_2 = 1$, and $h_1 = h_3 = h_4 = h_5 = 0$. The implied characteristic is $\mathbb{E}_0\left[S_T/S_0-1\right]$ and is the mean return.■

The *implied return* is zero. The *realized return* is the sum of daily net returns over the month. The corresponding *return swap* is a standard equity for floating swap – the receiver receives the total return on the underlying (net of the riskless interest rate) each "day" on a fixed nominal amount. The payer hedges by going long $1/S_i$ forward contracts each day.

3.3 Variance of Returns

Proposition 4: *the function* $g^V(\delta s) \equiv 2\left(e^{\delta s} - 1 - \delta s\right)$ *is a measure of variance and has the Aggregation Property.*

Proof: The Aggregation Property follows from Proposition 2 with h_1 = -2, h_2 = +2, $h_3 = h_4 = h_5 = 0$. $g^V(x) = x^2 + O\left(x^3\right)$, so g^V is a generalized variance measure. ■

With this unconventional definition of variance, the *implied variance* at time t, IV_t, is the price of a contract that pays $g^V(s_T\text{-}s_0)$, so $IV_t = v_t^L$. The *realized variance* is $RV_t = \sum^{\mathbf{T}} g^V(\delta s) = \sum^{\mathbf{T}} 2\left(e^{\delta s} - 1 - \delta s\right)$. The *true variance* is $TV_t = \mathbb{E}_t\left[2\left(S_T/S_0 - 1 - \ln\left(S_T/S_0\right)\right)\right]$ where the expectation is under the physical measure, rather than under a risk-adjusted measure. The true variance is unobservable; the realized variance is an unbiased estimate of the true variance if the price process is martingale; the implied variance is an unbiased estimate of the true variance in the absence of a variance risk premium.

The definition of implied variance is the same as the standard MFIV. The realized variance differs from the conventional definition, $\sum^{\mathbf{T}}(\delta s)^2$, but is found in Bondarenko (2010). The conventional definition does have the merit that it is the definition used in the variance swap market. But, as noted by Jiang and Tian (2005; see particularly footnote 7), the replication of a standard variance swap is imperfect. The replication is only perfect in the limit case when the mesh of the partition goes to zero, and with the added assumption that the price process is continuous. By contrast, the fact that the new measure of realized variance has the Aggregation Property means that replication is perfect for every price path and every partition, and is robust to jumps.

In practice, with reasonably frequent rebalancing, the two measures of realized variance are very similar. This is not surprising since they are both generalized measures of variance. The monthly realized volatility of the S&P500 computed using daily returns has averaged just over 18% (annualized) over the last ten years (2001-2010). The root mean square difference between the two measures over that period is 0.06%. Since the 1950s the biggest difference between the two measures was in the month of October 1987 when the conventional measure of realized volatility was 101.2%, while the alternative measure registered 98.8%.

There do not appear to be any strong theoretical arguments for preferring the conventional definition of realized variance (apart from the fact that it is well established both in the academic and practitioner communities). The main justification given in the literature for the conventional measure of realized variance is that it converges to the quadratic variation as the mesh size becomes small. The quadratic variation is important because it is an unbiased estimate of the conditional variance of the log price process under certain conditions (see Andersen, Bollerslev, Diebold and Labys, 2003, Theorem 1 and Corollary 1).[4] But, as the following Proposition states, *RV* too has this property when the process is a diffusion.

Proposition 5: *if f is an analytic function on the real line such that* $\lim_{x\to 0} f(x)/x^2 = 1$, *then for any continuous semimartingale s, the associated realized variance* $\sum^{\mathbf{T}} f(\delta s)$ *converges in probability to the quadratic variation of s as the mesh of the partition* **T** *goes to zero.*

Proof: see Appendix 1.[5]

In the empirical sections of this paper, I use the term variance in the sense of Proposition 4.

3.4 The Third Moment of Returns

Proposition 2 also shows how to construct a definition of the realized third moment of returns, one that closely resembles the definition already established for price changes.

Proposition 6: $g^Q\left(\delta s, \delta v^E\right) \equiv 3\delta v^E\left(e^{\delta s}-1\right)+K\left(\delta s\right)$ *, where* $K\left(\delta s\right) \equiv$ $6\left(\delta s e^{\delta s}-2e^{\delta s}+\delta s+2\right)$, *approximates the third moment of log returns and has the Aggregation Property.*

Proof: g^Q has the Aggregation Property by Proposition 2 with $h_1 = 6$, $h_2 = -12$, $h_3 = -3$, $h_4 = 0$, and $h_5 = 3$. $g^Q\left(s_T - s_t, v_t^E - v_T^E\right) = -3v_t^E\left(S_T/S_t - 1\right)+K\left(s_T - s_t\right)$. With the price following a martingale, the first term is zero in expectation. $K\left(x\right)=x^3+O\left(x^4\right)$ and converges to the third moment of returns when x is small. ■

The *implied third moment*, ITM_t, is the price of a claim that pays $g^Q\left(s_T - s_t, -v_t^E\right)$. It can be replicated exactly from forward contracts, entropy contracts and log contracts. The price of the claim is

$$ITM_t = 3\left(v_t^E - v_t^L\right). \tag{14}$$

In a Black-Scholes world, all options trade on the same implied volatility, the log and entropy variances are equal and $ITM_t = 0$. As Bakshi and Madan (2000) show, any general claim can be replicated by a portfolio of vanilla options; the replicating portfolio for the third moment claim is

$$6\left\{\int_{S_t}^{\infty}\frac{k-S_0}{S_t k^2}\mathbf{C}(k)dk-\int_0^{S_t}\frac{S_t-k}{S_t k^2}\mathbf{P}(k)dk\right\}-3v_t^E\mathbf{F}, \tag{15}$$

where $\mathbf{C}(k)$ and $\mathbf{P}(k)$ denote a call and a put with maturity T and strike k, and $\mathbf{F}$ is an at-the-money forward contract with the same maturity.

The implied third moment is thus the price of a portfolio that is long out of the money calls and short out of the money puts. By dividing the implied third moment by the implied variance to the power of 3/2, the implied skewness coefficient (ISC) can be computed

$$ISC_t = \frac{ITM_t}{\left(IV_t\right)^{3/2}}=\frac{3\left(v_t^E-v_t^L\right)}{\left(v_t^L\right)^{3/2}}. \tag{16}$$

The realized third moment is $RTM_t = \sum^{\mathbf{T}} g^Q\left(\delta s, \delta v^E\right)$; it is natural then to define the realized skewness coefficient

$$RSC_t = \frac{RTM_t}{\left(RV_t\right)^{3/2}} = \frac{\sum^{\mathbf{T}}\left\{3\delta v^E\left(e^{\delta s}-1\right)+K\left(\delta s\right)\right\}}{\left(\sum^{\mathbf{T}}2\left(e^{\delta s}-1-\delta s\right)\right)^{3/2}}. \tag{17}$$

Finally, the true third moment and true skew coefficient can be defined as $TTM_t = \mathbb{E}_t\left[g^Q\left(s_T - s_t, -v_t^E\right)\right]$ and $TSC_t = TTM_t\left(TV_t\right)^{-3/2}$.

The realized third moment is an unbiased estimator of the true third moment if prices of the underlying asset and of the entropy contract defined on it are martingales. The realized skew coefficient is not necessarily an unbiased estimate of the true skew coefficient since the ratio of two unbiased estimators is not in general an unbiased estimate of the ratio.

A third moment swap, where the floating leg is the realized third moment and the fixed leg is the implied third moment, can be replicated perfectly by dynamic hedging, trading the entropy contract and the forward contract. To replicate the swap, it is necessary that an entropy contract with maturity T is traded (or equivalently, that calls or puts with maturity T and all possible strikes are traded, so that the entropy contract can be replicated). An agent who writes the swap, receiving fixed and paying floating, receives net

$$3\left(v_0^E - v_0^L\right) - \sum^{\mathbf{T}}\left\{3\delta v^E\left(e^{\delta s} - 1\right) + K\left(\delta s\right)\right\}. \tag{18}$$

To hedge her position she needs to buy six log contracts, with terminal pay-off $6\ln S_T$. This costs $6\ln S_0 - 3v_0^L$ (by (10)). She also needs to hedge dynamically by holding a long position of $6/S_i$ entropy contracts on day i (contracts that have a pay-off of $S_T \ln S_T$) and a short position in $3\left(2s_i + v_i^E + 4\right)/S_i$ forward contracts. The replication of the swap is exact.

3.5 True and Realized Third Moments when Prices are not Martingales

The realized third moment is useful for estimating the true third moment of returns. Proposition 6 shows that the realized third moment is an unbiased estimator of the true third moment when prices are martingale. But when prices are not martingale, it is a biased estimator. The following proposition characterizes the bias under the much weaker assumption that the process is ergodic.

Proposition 7 *The difference between the true and realized third moments depends on the cross-correlation between returns on the underlying asset and the returns on a hedged option position. Specifically*

$$TTM = \mathbb{E}\left[RTM\right] + 6\mathbb{E}\left[\sum_{u=1}^{T-1}\sum_{t=u}^{T-1}\left\{\frac{\delta E_u - \Delta_0 \delta S_u}{S_0}\frac{\delta S_{t+1}}{S_t} + \frac{\delta S_u}{S_0}\frac{\delta E_{t+1} - \Delta_t \delta S_{t+1}}{S_t}\right\}\right] \tag{19}$$

where E_t is the price of the entropy contract at time *t*, and $\Delta_t \equiv \dfrac{\partial E_t}{\partial S_t} = 1 + \ln S_t + \frac{1}{2} v_t^E$.

If changes in the term structure of volatility are parallel, and the correlation between returns on the underlying asset on day t and returns on a fixed maturity delta-hedged entropy contract on day t+n is $\rho_{ry}(n)$ then

$$\frac{TTM}{\mathbb{E}\left[RTM\right]} \approx \frac{\sum_{n=-T}^{T} w_{TM}\left(n,T\right)\rho_{ry}\left(n\right)}{\rho_{ry}\left(0\right)} \text{ where } w_{TM}\left(n,T\right) \equiv \frac{\left(T-|n|\right)\left(T-n-1\right)}{T\left(T-1\right)}. \tag{20}$$

The difference between the true variance of a price series and its realized variance depends on the auto-correlation function of returns, $\rho_{rr}\left(u\right)$.

$$\frac{TV}{\mathbb{E}\left[RV\right]} \approx \sum_{n=-T}^{T} w_V\left(n,T\right)\rho_{rr}\left(n\right) \text{ where } w_v\left(n,T\right) \equiv \frac{\left(T-|n|\right)}{T}, \tag{21}$$

Proof: *in Appendix* 1.

The first part of the proposition says that if the expected return on an asset is positively (negatively) correlated with its variance risk premium in prior or in subsequent periods, then the realized third moment is a downward (upward) biased estimate of the true third moment. Under the hypothesis that the underlying asset and the entropy contract are martingales, the expected return on the asset and its variance risk premium are zero, and hence there is no bias.

The second part expresses the bias in terms of the cross-correlation between the expected return and the variance premium. The assumption that shifts in term structure are parallel allows the return on an option contract with a maturity that declines in time to be replaced by the return on a duration matched constant maturity option.

The third part, concerning the relation between realized and true variance is well-known in the literature (see Campbell, Lo and MacKinlay (1996) page 49). The relationship between realized variance estimates computed using different horizons has long been used as a test for serial correlation in returns (Lo and MacKinlay, 1988).

4. Simulations

As seen in section 1, direct estimates of the skewness of monthly returns are noisy, even with many years of data. Realized skewness, which makes use of higher frequency data, offers more precise estimates. In the presence of equity and variance risk premia, Proposition 7 shows that the estimates are biased. In this section, simulation is used to quantify both the improvement in precision and the bias.

The process that is simulated is the SVCJ stochastic volatility model of Duffie, Pan and Singleton (2000) with contemporaneous jumps in the underlying and volatility

$$\begin{aligned} \frac{dS_t}{S_t} &= \gamma\, dt + \sqrt{V_t}\, dW_t^S + \left(e^{Z_t^S} - 1\right) dN_t \\ dV_t &= \kappa\left(\theta - V_t\right) dt + \sigma_V \sqrt{V_t}\, dW_t^V + Z_t^V dN_t \end{aligned} \tag{22}$$

S is the price of the underlying, V is its spot variance, W^S and W^V are Brownian processes with correlation ρ, N is a Poisson process with intensity λ, Z^S is normally distributed with mean μ_S and standard deviation σ_S, while Z^V is distributed exponentially with mean μ_V. Expressions for the unconditional true, implied and realized moments of τ-horizon returns under the SVCJ model can be obtained analytically. They are set out in Appendix 2.

The parameters used for the simulation, shown in Table 1, are taken from Broadie, Chernov and Johannes (BCJ) (2007; Table I "EJP", and Table IV). They are estimated from S&P futures and options data from 1987 to 2003. The drift γ is chosen so that S is martingale under the physical measure as well as the risk-adjusted measure. As a result, the realized variance is an unbiased estimate of the true variance. The difference in the parameters of the volatility process under the two measures implies the existence of variance risk premia that create a bias in the realized third moment. The risk-adjusted parameters are used to compute option prices and the implied log and entropy variances. Paths for the asset price and implied variances are simulated[6] under the physical measure, using daily increments over 200 months, each of 22 days. For each run, the distribution of the 200 monthly returns is used to compute the sample second and third moments as well as the skewness coefficient. The monthly realised and implied moments are calculated and averaged over the 200 months.

The top panel of Table 2 shows that the sample variance of monthly returns computed over 200 months provides a fairly precise and unbiased estimate of the actual variance. The unconditional variance of monthly returns implied by the parameters in Table 1 is 0.214×10^{-2}; the sample estimate has a standard deviation of 0.038×10^{-2}. The estimate of the third moment also appears to have little bias, but has much greater noise. The standard deviation of the third moment is of similar magnitude to its true value; 200 months is too short a period to reject the hypothesis that monthly returns are positively skewed.

The implied moments have far smaller standard deviations than their sample counterparts; this is to be expected since they are beginning-of-month expectations rather than end-of-month out-turns. They are strongly biased because of the risk premia in option prices, reflected in the large difference between the physical and risk-adjusted parameters in Table 1.

The realized second moment is an unbiased estimate of the unconditional second moment. This is expected since the simulated price process is martingale. It is also less noisy than the sample second moment. The realized third moment is also much less noisy than its sample counterpart, showing its value as a tool for estimating skewness. It does exhibit substantial bias. With the implied variance being larger than the expected variance one might expect the realized third moment (which is related to the covariation between implied variance and returns) to be biased upwards in absolute terms. This is not the case. As shown in Appendix 2, the bias in the realized third moment is due entirely to the difference in the speed of mean reversion (κ) of volatility under the two measures. Mean reversion is twice as large under the risk adjusted measure as under the physical measure. This reduces the beta of implied variance on instantaneous volatility, and so reduces the magnitude of the covariation between implied variance and returns.

BCJ's estimate of the difference in the speed of mean reversion under the two measures ($\kappa^Q - \kappa^P$) is driven by the estimated diffusive risk premium (which they call η_V). As they note (BCJ, 2007, p 1476) it is hard to estimate the sign of the diffusive risk premium let alone its magnitude. While it is clear that this premium may give rise to significant bias, the sign of the bias that arises in practice cannot be identified with confidence.

The implications for the estimates of the skewness coefficient follow immediately from the estimates of the second and third moments. The sample estimate is very noisy, with a standard deviation almost equal to its mean. Although both moment estimates are unbiased, the sample skewness coefficient is biased. The implied skewness coefficient is much more precise, but is more than three times too large, while the realized estimate is also biased being only two thirds of the true value.

The second panel is similar to the first except that it is concerned with annual returns estimated using 20 years of data. The true unconditional skewness of annual returns is somewhat lower than of monthly returns, but with only 20 rather than 200 observations, the sample estimate is even noisier. The implied skewness is still biased, but the bias is somewhat smaller than in the monthly data. The magnitude of the bias in the realized skewness is much greater. At longer horizons, the increased speed of mean reversion in volatility greatly attenuates the covariation between returns and implied variance.

The main conclusions that can be drawn from the simulations is that realized skewness provides a far more precise estimate of skewness than does the sample skewness, but it is subject to bias if the risk neutral speed of mean reversion in volatility differs from its physical counterpart. Variance risk premia that do not affect the covariation between returns and implied volatility (such as the variance jump risk premia in the SVCJ model) have no effect on realized skewness though they do give rise to bias in the implied skewness.

5. Skewness of Equity Index Returns

5.1 The Term Structure of Skewness

The empirical analysis in this section is based on European options written on the S&P500 index traded on the CBOE obtained from OptionMetrics. The one month options mature every month, while the 3, 6 and 12 month options mature every three months. In each of the series, the first period starts in December 1997 and the last starts in September 2009. The data set includes closing bid and ask quotes for each option contract along with the corresponding strike prices, Black-Scholes implied volatilities, the zero-yield curve, and closing spot prices of the underlying. Entries with non-standard settlements are deleted.

Put and call option prices for every strike at each maturity are computed by interpolating implied volatilities between quoted strike prices using a cubic spline. Outside the quoted range, the implied volatilities at the lowest and the highest strike price are used. The Log and Entropy Contracts are synthesized from the continuum of conventional options and v^E and v^L are calculated. This is the procedure used by Carr and Wu (2009).

Table 3 shows summary statistics at different maturities. Realized variance is on average lower than implied variance, which is consistent with a positive variance risk premium, and both increase linearly with maturity as one might expect. The realized and implied third moments are negative at all maturities[7]. They increase with maturity faster than linearly. The implied third moment is on average larger (in absolute size) than the realized third moment at short maturities, but the difference appears to vanish or reverse at longer maturities. The skewness, both implied and realized, exceeds -1 on average at all maturities. Implied skewness appears to decline with maturity, whereas realized skewness on average actually increases with maturity.

Both the second and third moments, whether realized or implied, are highly skewed and variable. They are very highly (negatively) correlated with each other, with correlations in excess of -0.9. The skewness coefficients tend to be much less variable and to be distributed symmetrically. They are much less highly correlated with variance; the sign of the correlation coefficient is positive implying that the higher the variance the less skewed the distribution, but the correlations are all below 0.5.

To analyze the term structure of skewness, define $Y_{t,n}$ as the realized third moment accumulated over the quarter starting at time t using options that mature at time $t+n$ where time is measured in months. The Y are all negative, and are highly correlated in the cross-section, with the magnitude increasing with maturity. Table 4 reports the results of the regression

$$\log\left(Y_{t,n}/Y_{t,3}\right)=\sum_{i=6,9,12}\log\left(\beta_i\right)D_i+\tilde{\varepsilon}_{t+3,n}\,. \tag{23}$$

where D_i is a dummy that takes the value of 1 if $n = i$, and zero otherwise, and $\tilde{\varepsilon}$ is an error term. The estimated value of β_6 is 2.17, so the realized skewness over a quarter using options that expire three months after the end of the quarter is 2.17 times as high as the realized skewness computed using options that expire at the end of the quarter. This implies that the realized third moment over six months is 3.17 its value over 3 months. Realized variance is linear in horizon, so the realized skewness coefficient over six months is $3.17/2^{1.5} = 1.12$ times its value over three months. The increase is statistically significant; $\log\beta_3$ would need to be below 0.60 for the skewness to be lower at six months than at three, and $t(\log\beta_3\text{-}0.60) = 3.21$.

The point estimate of relative skewness increases from 1 at three months to 1.12 at six months, and to 1.15 at nine and twelve months, though the increases beyond six months are not statistically significant. The panel regression of Table 4 therefore confirms the impression given by the population averages of Table 3 that realized skewness rises with horizon up to six months, and shows no evidence of decline in horizons up to one year.

Before concluding that true skewness too increases with horizon, it is necessary to address the issue of bias. Proposition 7 shows how cross correlations between index returns and returns on a hedged constant maturity entropy contract can be used to estimate the size of any bias. To compute the option returns, I price a notional entropy contract whose maturity is set equal to two-thirds of the horizon of interest[8] by interpolating linearly between the entropy contracts with neighboring maturities. To put some error bounds round the point estimates, I also compute the same statistics by bootstrapping the two series, destroying their auto-correlation structure. The results are set out in Table 5.

This shows that the realized second and third moments both over-estimate the true moments at all maturities, with the effect being more important (and more significant statistically) for the variance than for the third moment, and for shorter horizons than for longer horizons. It is not possible to draw firm conclusions about bias in the skewness coefficient since it is the ratio of two biased estimates, but at the very least one can say that the data do not suggest that the realized skewness over-estimates the true skewness, and they also suggest that any bias in the moment estimates attenuates with maturity.

The analysis of realized skewness of index returns at horizons of up to one year shows that returns are strongly negatively skewed at all horizons (with a skew coefficient in excess of -1), that there is no evidence at all that skewness declines with horizon and some evidence that it actually increases for horizons up to six months.

5.2 Time Variation in Skewness

Realized skewness is quite volatile. Table 3 shows that the realized skewness of the S&P500 on a quarterly horizon has averaged -1.39 with a standard deviation of 0.5. The following forecasting model is used to investigate whether this variation is predictable or whether it is just noise

$$\begin{aligned} &RSC_{t,t+n} = \alpha + \beta_1 ISC_{t,t+n} + \beta_2 RSC_{t-m,t} + \varepsilon_{t+n} \\ &\text{for } n = 1,3,6 \text{ and } 12, \text{ and } m = \min\{3,n\}. \end{aligned} \tag{24}$$

Here $RSC_{t,t+n}$, the realized skew from time t to $t+n$, is forecast using $ISC_{t,t+n}$ the skew implied by the time t prices of options that expire at $t+n$, and the lagged realized skew. The results in Table 6 show that there is predictable time variation in the realized skew, with between 10% and 25% of the variation in the realized skew being predictable using simple explanatory variables. The lagged realized skew enters in with positive sign for shorter maturities suggesting that the skew risk premium is predictable, but the magnitude is small. The coefficient is not statistically significant at longer horizons.

Figure 3 shows the relationship between realized and predicted skew graphically at the quarterly horizon, and demonstrates significant time variation in the skew, with predicted skew varying between -1.0 and -1.8 over the period. Interestingly, the period when index volatility was very low by historic standards (2003-7) was also one of relatively high skewness, while in the volatility spike of 2008 skewness was actually rather low.

5.3 The Economic Significance of Index Skewness

To quantify the economic significance of the skewness of index returns, I compute the Markowitz risk premium (expressed as a rate of return) that a rational agent with constant relative risk aversion γ would demand to compensate for the risk of investing 100% in the market rather than 100% in the riskless asset over a particular horizon τ. Let r denote the excess log return on the asset. In the classic case of an asset with constant volatility σ and zero drift (Merton, 1969), the premium required is $\mu = \frac{1}{2}\gamma\sigma^2$. When r is not lognormally distributed, the premium depends also on higher moments (as noted for example by Kraus and Litzenberger, 1976), and the premium rate required is

$$\mu = -\frac{\log \mathbb{E}\left[e^{\tilde{r}(1-\gamma)}\right]}{(1-\gamma)\tau}. \tag{25}$$

I assume that the moment generating function for r takes the form

$$m_r(t) \equiv \mathbb{E}\left[e^{t\tilde{r}}\right] = e^{f(t)} \tag{26}$$

for some function *f*. Then

$$\mu = -\frac{f(1-\gamma)}{(1-\gamma)\tau}. \tag{27}$$

f is calibrated to meet four conditions: the probability density integrates to 1, the price is martingale, and the distribution matches the log and entropy variances

$$\begin{aligned}
&\mathbb{E}[1] = 1 \Rightarrow e^{f(0)} = 1; \\
&\mathbb{E}\left[e^r\right] = 1 \Rightarrow e^{f(1)} = 1; \\
&\mathbb{E}[r] = -v^L/2 \Rightarrow f'(0)e^{f(0)} = -v^L/2; \\
&\mathbb{E}\left[re^r\right] = -v^E/2 \Rightarrow f'(1)e^{f(1)} = v^E/2.
\end{aligned} \tag{28}$$

To evaluate the function $f(t)$ at t = 1-γ I extrapolate using a cubic polynomial that fits the four conditions in (28)[9]. This allows the premium to be expressed as a function of the variance (*TV*) and the third moment (*TTM*) of the return r

$$\begin{aligned}
\mu &= -\frac{f(1-\gamma)}{(1-\gamma)\tau} \\
&\approx \frac{\gamma}{2\tau}\left(v^L - (\gamma-1)\left(v^E - v^L\right)\right) \\
&= \frac{\gamma}{2\tau}\left(TV - \frac{\gamma-1}{3}TTM\right).
\end{aligned} \tag{29}$$

In the absence of skew, the third moment is zero and this reverts to the Merton result. Any investor with γ > 1 has a preference for positive skewness, and the skewness risk premium (positive if skewness is negative) is proportional to the third moment of returns as well as to the agent's risk aversion coefficient.

Using the mean observed realized variance and third moment of the S&P500 over 1997-2009 from Table 3 as estimates of the true variance and third moment, this implies that for an investor with $\gamma = 2$, the required risk premium μ is 4.83%/year at the monthly horizon; had the distribution been unskewed (with $v^E = v^L$) the premium would have been 4.68%, a difference of 15 basis points annually. At longer horizons, the skew risk premium is higher. For example, for an investor with a one year horizon, the required risk premium is 5.35%/year, against 4.75% for a symmetrical distribution – a skew risk premium of 60 basis points annually.

At higher level of risk aversion, the risk premium is also higher, and the skew risk premium is more important both absolutely and relatively. So with $\gamma = 3$, the skew risk premium at one year contributes 181 basis points to a total risk premium of 8.94%.

This analysis suggests that the levels of skewness observable in US stock market returns, particularly at long horizons, and the predictable variability in the levels of skewness documented in this paper are economically as well as statistically significant for investors with quite moderate degrees of risk aversion.

6. Conclusions

Skewness of returns at long horizons is important for both asset pricing and risk management, yet measuring skewness is very difficult. With almost fifty years of data the evidence that daily index returns are negatively skewed just meets standard significance levels, but even then the results are heavily influenced by the Crash of 1987. For longer horizon returns, there is also evidence that returns are negatively skewed. But the confidence intervals are wide, and the results are insignificant in many ten year sub-periods. There is however strong evidence that the distribution of long horizon returns can only be explained by serial dependencies in the data, and is inconsistent with an iid model of the returns process. There is a need for a methodology to measure the skewness of long horizon returns that is far more precise than simply using the sample skewness, and which respects the serial dependencies.

In the case of the second moment of returns, the problem has been addressed by using high frequency returns to compute the realized variance over long horizons. The key property of realized variance is that it is an unbiased estimate of the true conditional variance. In the paper this property has been formalized as the Aggregation Property. The standard definition of realized variance needs to be slightly modified so that it too has the Aggregation Property. More importantly, there is a unique definition of the realized (and true) third moment under which the realized third moment is an unbiased estimate of the true third moment. Calculating the realized third moment requires data on option price returns as well as the returns on the underlying.

The realized second and third moments of index returns are highly correlated, but there is significant variation in the realized skewness coefficient and it is predictable; implied skewness from option prices (as measured by the slope of implied volatility with strike) predicts realized skewness (which roughly corresponds to the correlation between returns and volatility shocks). The realized skewness of the market index actually increases with the horizon, at least from one to twelve months. This is inconsistent with an iid model in which the twelve month skewness would be less than 30% of the one month skewness. The degree of skewness in returns is significant economically as well statistically, substantially increasing the risk premium required by a rational investor for holding the market. This is particularly marked for investors with long horizons.

There are many further research questions that could usefully be explored with the use of realized skewness. The examination of the relationship between implied variance and realized variance, and the existence of variance risk premia by Carr and Wu (2009) could readily be extended to skewness risk premia both in the equity index market (as in Kozhan, Neuberger and Schneider (2011)) and in other financial markets. The evidence on the pricing of skewness and coskewness presented by Ang, Hodrick, Xing and Zhang (2006) could be refined using a definition of skewness with a more rigourous theoretical basis.

It would also be nice to be able to extend the analysis to higher order moments (or indeed to the cumulants of the distribution as in Backus, Chernov and Martin (2011)). This would not be straightforward; as Proposition 2 shows, the set of functions that possess the Aggregation Property is quite limited; the way forward here may be to include other traded claims in addition to those on the variance of the distribution.

APPENDIX 1 - PROOFS

Proof of Proposition 1

It is straightforward to prove that all members of **G** have the Aggregation Property; all that is needed is to substitute (3) into (2). Proving the converse, that all analytic functions that have the Aggregation Property are in **G**, is more complicated.

Let $\tilde{\eta}$ be a random variable with $\mathbb{E}\left[\tilde{\eta}\right]=0$ and $\mathbb{E}\left[\tilde{\eta}^2\right]=\alpha$, and g an analytic function that has the Aggregation Property. Consider two processes with $t\in\{0,1,2\}$. The first is given by $S_0 = S_1 = 0$, $S_2 = \tilde{\eta}$. S is clearly martingale. The process (S, V) is (0, α), (0, α), ($\tilde{\eta}$, 0). For the Aggregation Property to hold

$$\mathbb{E}\left[g\left(\tilde{\eta},-\alpha\right)\right]=\mathbb{E}\left[g\left(0,0\right)+g\left(\tilde{\eta},-\alpha\right)\right]. \tag{30}$$

It follows that $g(0, 0) = 0$. The second process for S is

$$0\rightarrow\begin{cases} u & \rightarrow \quad u+\tilde{\eta} \qquad \Pr=\pi \\ d & \rightarrow \quad d \qquad \Pr=1-\pi \end{cases}, \text{ with } ud<0 \text{ and } \pi u+\left(1-\pi\right)d=0. \tag{31}$$

S is martingale. The process (S, V) is

$$\begin{gathered}\left(0,V_0\right)\rightarrow\begin{cases} \left(u,\alpha\right) & \rightarrow \quad \left(u+\tilde{\eta},0\right) \qquad \Pr=\pi \\ \left(d,0\right) & \rightarrow \quad \left(d,0\right) \qquad \Pr=1-\pi \end{cases} \\ \text{where } V_0=\pi\left(u^2+\alpha^2\right)+\left(1-\pi\right)d^2. \end{gathered} \tag{32}$$

For the Aggregation Property to hold

$$\begin{aligned}&\mathbb{E}\left[\pi g\left(u+\tilde{\eta},-V_0\right)+\left(1-\pi\right)g\left(d,-V_0\right)\right]=\\&\mathbb{E}\left[\pi g\left(u,\alpha-V_0\right)+\pi g\left(\tilde{\eta},-\alpha\right)+\left(1-\pi\right)g\left(d,-V_0\right)+\left(1-\pi\right)g\left(0,0\right)\right].\end{aligned} \tag{33}$$

Simplifying, and using the fact that $g(0, 0) = 0$, gives

$$\mathbb{E}\left[g\left(u+\tilde{\eta},-V_0\right)\right]=\mathbb{E}\left[g\left(\tilde{\eta},-\alpha\right)\right]+g\left(u,\alpha-V_0\right), \tag{34}$$

for arbitrary u and V_0. Take the limit of (34) as $u\to 0$

$$\mathbb{E}\left[g\left(\tilde{\eta},-V_0\right)\right]=\mathbb{E}\left[g\left(\tilde{\eta},-\alpha\right)\right]+g\left(0,\alpha-V_0\right). \tag{35}$$

Take the derivative of (35) with respect to V_0

$$\mathbb{E}\left[g_2\left(\tilde{\eta},-V_0\right)\right]=g_2\left(0,\alpha-V_0\right), \tag{36}$$

where the subscript denotes the partial derivative. Now take limits as $V_0\to\alpha$

$$\mathbb{E}\left[g_2\left(\tilde{\eta},-\alpha\right)\right]=g_2\left(0,0\right), \tag{37}$$

(37) holds for any random variable $\tilde{\eta}$ with $\mathbb{E}\left[\tilde{\eta}\right]=0$ and $\mathbb{E}\left[\tilde{\eta}^2\right]=\alpha$. So for any positive function p

$$\begin{aligned}&\int_{-\infty}^{+\infty}p\left(S\right)g_2\left(S,V\right)dS \text{ is constant provided that } \int_{-\infty}^{+\infty}p\left(S\right)dS=1,\\&\int_{-\infty}^{+\infty}Sp\left(S\right)dS=0 \text{ and } \int_{-\infty}^{+\infty}\left(S^2+V\right)p\left(S\right)dS=0.\end{aligned} \tag{38}$$

The Lagrangian of system (38), $g_2\left(S,V\right)-\lambda_1-\lambda_2 S-\lambda_3\left(S^2+V\right)$, is zero, so $g_2(S, V)$ must take the form

$$g_2(S,V) = a + B(V)S + C(V)(S^2 + V), \tag{39}$$

for some constant a and functions B and C. But substituting (39) back into (36) shows that $C(v)$ is a constant, denoted by $2c$. Integrating (39) gives

$$g(S,V) = aV + S\int_0^V B(W)dW + c(2S^2 + V)V + D(S), \tag{40}$$

where D again is an arbitrary function. It is easy to verify that (40) does indeed satisfy (35) provided that $D(0) = 0$. Substituting it into the more general (34) shows that $c = 0$, and that the following must be satisfied if g is to have the Aggregation Property

$$u\int_{\alpha - V_0}^{-V_0} B(W)dW + \mathbb{E}\left[D(u+\tilde{\eta}) - D(\tilde{\eta}) - D(u)\right] = 0, \tag{41}$$

for arbitrary u, V_0 and random variable $\tilde{\eta}$ with zero mean. For this to hold, differentiating (41) with respect to V_0 gives $B(\alpha - V_0) = B(-V_0)$, so $B(V)$ must be a constant, denoted by $3b$. Let

$$\tilde{\eta} = \tilde{\eta}^*(\kappa) \equiv \begin{cases} +\sqrt{\kappa} & \Pr = 1/2 \\ -\sqrt{\kappa} & \Pr = 1/2 \end{cases} \tag{42}$$

for some $\kappa \in (0,1)$. Substitute into (41), divide by $\kappa/2$ and take limits as $\kappa \to 0$

$$\begin{aligned} D''(u) - D''(0) &= 6bu, \text{ so} \\ D(u) &= d_1 u + d_2 u^2 + bu^3, \end{aligned} \tag{43}$$

for any d_1, and d_2. So g must take the form

$$g\left(S,V\right)=h_0V+h_1S+h_2S^2+h_3\left(S^3+3SV\right) \tag{44}$$
where $h_0=a, h_1=d_1, h_2=d_2$ and $h_3=b$.

. ■

Proof of Proposition 2

The proof is similar to the proof of Proposition 1, but with the added problem that the form of the variance function is not known. The proof that all members of **G*** have the Aggregation Property is straightforward. This proof focuses on the converse.

Let $\tilde{\eta}$ be a random variable with $\mathbb{E}\left[e^{\tilde{\eta}}\right]=1$ and $\mathbb{E}\left[f\left(\tilde{\eta}\right)\right]=\alpha$, and g a function that has the Aggregation Property. Consider two processes with $t\in\{0,1,2\}$. The first is given by $s_0 = s_1 = 0$, $s_2 = \eta$. $S = e^s$ is clearly martingale. The process (s, v) is $(0, \alpha)$, $(0, \alpha)$, $(\eta, 0)$. For the Aggregation Property to hold

$$\mathbb{E}\left[g\left(\tilde{\eta},-\alpha\right)\right]=\mathbb{E}\left[g\left(0,0\right)+g\left(\tilde{\eta},-\alpha\right)\right]. \tag{45}$$

It follows that $g(0, 0) = 0$. The second process for s is

$$0\rightarrow\begin{cases} u & \rightarrow \quad u+\tilde{\eta} \qquad \Pr=\pi \\ d & \rightarrow \quad d \qquad \Pr=1-\pi \end{cases}, \text{with } ud<0 \text{ and } \pi e^u+\left(1-\pi\right)e^d=1. \tag{46}$$

S is martingale. The process (s, v) is

$$\left(0,v_0\right)\rightarrow\begin{cases} \left(u,\alpha\right) & \rightarrow \quad \left(u+\tilde{\eta},0\right) \qquad \Pr=\pi \\ \left(d,0\right) & \rightarrow \quad \left(d,0\right) \qquad \Pr=1-\pi \end{cases} \tag{47}$$
where $v_0=\pi\mathbb{E}\left[f\left(u+\tilde{\eta}\right)\right]+\left(1-\pi\right)f\left(d\right)$.

For the Aggregation Property to hold

$$\begin{gathered}\mathbb{E}\left[\pi g\left(u+\tilde{\eta},-v_0\right)+(1-\pi) g\left(d,-v_0\right)\right]= \\ \mathbb{E}\left[\pi g\left(u, \alpha-v_0\right)+\pi g(\tilde{\eta},-\alpha)+(1-\pi) g\left(d,-v_0\right)+(1-\pi) g(0,0)\right].\end{gathered} \tag{48}$$

Simplify, and use the fact that $g(0, 0) = 0$, to give

$$\mathbb{E}\left[g\left(u+\tilde{\eta},-v_0\right)\right]=\mathbb{E}[g(\tilde{\eta},-\alpha)]+g\left(u, \alpha-v_0\right), \tag{49}$$

for arbitrary u and $v_0 > 0$. Take the limit of (49) as $u \to 0$

$$\mathbb{E}\left[g\left(\tilde{\eta},-v_0\right)\right]=\mathbb{E}[g(\tilde{\eta},-\alpha)]+g\left(0, \alpha-v_0\right), \tag{50}$$

Take the derivative of (50) with respect to v_0

$$\mathbb{E}\left[g_2\left(\tilde{\eta},-v_0\right)\right]=g_2\left(0, \alpha-v_0\right), \tag{51}$$

where the subscript denotes the partial derivative. Now take limits as $v_0 \to \alpha$

$$\mathbb{E}\left[g_2(\tilde{\eta},-\alpha)\right]=g_2(0,0), \tag{52}$$

Since (52) holds for any random variable $\tilde{\eta}$ with $\mathbb{E}\left[e^{\tilde{\eta}}\right]=1$ and $\mathbb{E}[f(\tilde{\eta})]=\alpha$, using the same Lagrangian argument as in (38), $g_2(s, v)$ must take the form

$$g_2(s, v)=a+B(v)\left(e^s-1\right)+C(v)(f(s)+v), \tag{53}$$

for some constant a and functions B and C. Substituting (53) back into (51) shows that $C(v)$ is a constant, which is denoted by $2c$. Integrating (53) gives

$$g\left(s,v\right)=av+\left(e^{s}-1\right)\int_{0}^{v}B\left(w\right)dw+cv\left(2f\left(s\right)+v\right)+D\left(s\right), \tag{54}$$

where D again is an arbitrary function. It is easy to verify that (54) does indeed satisfy (50). Substituting it into the more general (49) shows that the following equation must be satisfied if g is to have the Aggregation Property

$$\begin{aligned}\left(e^{u}-1\right)\int_{\alpha-v_0}^{-v_0}B\left(w\right)dw-2c\left\{v_0\mathbb{E}\left[f\left(u+\tilde{\eta}\right)-f\left(u\right)\right]+\left(f\left(u\right)-v_0\right)\mathbb{E}\left[f\left(\tilde{\eta}\right)\right]\right\}\\ +\mathbb{E}\left[D\left(u+\tilde{\eta}\right)-D\left(\tilde{\eta}\right)-D\left(u\right)\right]=0\end{aligned} \tag{55}$$

For a random variable $\tilde{\eta}$ and $p\in\left[0,1\right]$ define

$$\tilde{\eta}_p\equiv\begin{cases}\tilde{\eta} & \Pr=p\\ 0 & \Pr=1-p\end{cases}. \tag{56}$$

If $\mathbb{E}\left[e^{\tilde{\eta}}\right]=1$ and $\mathbb{E}\left[f\left(\tilde{\eta}\right)\right]=\alpha$ then $\mathbb{E}\left[e^{\tilde{\eta}_p}\right]=1$ and $\mathbb{E}\left[f\left(\tilde{\eta}_p\right)\right]=\alpha p$. Putting $\tilde{\eta}_p$ into (55) gives

$$\begin{aligned}\left(e^{u}-1\right)\int_{p\alpha-v_0}^{-v_0}B\left(w\right)dw-2cp\left\{v_0\mathbb{E}\left[f\left(u+\tilde{\eta}\right)-f\left(u\right)\right]+\left(f\left(u\right)-v_0\right)\mathbb{E}\left[f\left(\tilde{\eta}\right)\right]\right\}\\ +p\mathbb{E}\left[D\left(u+\tilde{\eta}\right)-D\left(u\right)-D\left(\tilde{\eta}\right)\right]-\left(1-p\right)D\left(0\right)=0.\end{aligned} \tag{57}$$

By setting $p = 0$, it can be seen that $D(0) = 0$. Since the other terms in (57) are linear in the arbitrary scalar p , the first term must be so too, which implies that B is constant, and can be denoted by $-b$. So (55) can be simplified to

$$\begin{aligned}b\left(e^{u}-1\right)\mathbb{E}\left[f\left(\tilde{\eta}\right)\right]-2c\left\{v_0\mathbb{E}\left[f\left(u+\tilde{\eta}\right)-f\left(u\right)\right]+\left(f\left(u\right)-v_0\right)\mathbb{E}\left[f\left(\tilde{\eta}\right)\right]\right\}\\ +\mathbb{E}\left[D\left(u+\tilde{\eta}\right)-D\left(\tilde{\eta}\right)-D\left(u\right)\right]=0\end{aligned} \tag{58}$$

Let

$$\tilde{\eta} = \tilde{\eta}^*(\kappa) \equiv \begin{cases} \ln\left(1+\sqrt{\kappa}\right) & \Pr = 1/2 \\ \ln\left(1-\sqrt{\kappa}\right) & \Pr = 1/2 \end{cases} \tag{59}$$

for some $\kappa \in (0,1)$. Substitute into (58), divide by $\kappa/2$ and take limits as $\kappa \to 0$

$$\begin{gathered} 2b\left(e^u - 1\right) - 2c\left\{v_0\left(f''(u) - f'(u) - 2\right) + 2f(u)\right\} \\ +\left(D''(u) - D'(u) - D''(0) + D'(0)\right) = 0 \end{gathered} \tag{60}$$

Since (60) holds for arbitrary v_0

$$c = 0, \text{ or } f''(u) - f'(u) - 2 = 0. \tag{61}$$

If $c \neq 0$, solve for f using its limit properties at 0 to give

$$f(u) = 2\left(e^u - 1 - u\right) = L(u). \tag{62}$$

The general solution for D from (60) is

$$D(u) = d_1 u + d_2\left(e^u - 1\right) + \left(8c - 2b\right)ue^u + 4cu^2 \tag{63}$$

where d_1 and d_2 are arbitrary scalars. Putting (63) into (54) gives

$$\begin{gathered} g(s,v) = h_1 s + h_2\left(e^s - 1\right) + h_3 v + h_4\left(v - 2s\right)^2 + h_5 e^s\left(v + 2s\right), \text{ where} \\ h_1 = d_1; \quad h_2 = d_2; \quad h_3 = a + b - 4c; \quad h_4 = c; \quad h_5 = 4c - b. \end{gathered} \tag{64}$$

Finally, substituting for g into (49) gives

$$
\begin{aligned}
&\mathbb{E}\left[g\left(u+\tilde{\eta},-v_0\right)\right]-\mathbb{E}\left[g\left(\tilde{\eta},-\alpha\right)\right]-g\left(u,\alpha-v_0\right)= \\
&h_4\left(4u+2v_0-2\alpha\right)\left(\mathbb{E}\left[2\tilde{\eta}\right]+\alpha\right)+h_5\left(e^u-1\right)\left(\mathbb{E}\left[2\tilde{\eta}e^{\tilde{\eta}}\right]-\alpha\right).
\end{aligned} \tag{65}
$$

For this to be zero, as required for Aggregation, one of three conditions is necessary

$$
\begin{aligned}
&1)\quad h_4=h_5=0; \\
&2)\quad h_4=0,\ \text{and } \mathbb{E}\left[f\left(\tilde{\eta}\right)\right]=\mathbb{E}\left[2\tilde{\eta}e^{\tilde{\eta}}\right], \\
&\qquad \text{so } f\left(x\right)=2\left(xe^x-e^x+1\right)=E\left(x\right); \\
&3)\quad h_5=0,\ \text{and } \mathbb{E}\left[f\left(\tilde{\eta}\right)\right]=\mathbb{E}\left[-2\tilde{\eta}\right], \\
&\qquad \text{so } f\left(x\right)=2\left(e^x-1-x\right)=L\left(x\right).
\end{aligned} \tag{66}
$$

■

Proof of Proposition 5

By assumption, *f* is an analytic function which can be written as

$$
\begin{aligned}
&f\left(x\right)=x^2+h\left(x\right), \\
&\text{where } h\left(x\right)=\sum_{k\geq 3}a_k x^k.
\end{aligned} \tag{67}
$$

and the convergence radius of *f* is infinity, so

$$
\limsup_{k\to\infty}\sqrt[k]{\left|a_k\right|}=0. \tag{68}
$$

Let X be a continuous semimartingale. Assume, without loss of generality, that $X(0) = 0$. For a function g, define the *g-variation* of X as

$$
\begin{aligned}
&V\left(X,g\right)_t\equiv\lim_{n\to\infty}V^{(n)}\left(g\right)_t\ \text{ (if it exists), where} \\
&V^{(n)}\left(X,g\right)_t\equiv\sum_{i=1}^{[nt]}g\left(X_{i/n}-X_{(i-1)/n}\right),
\end{aligned} \tag{69}
$$

where $[x]$ denotes the integer part of x. Let T_N be the following increasing sequence of stopping times

$$T_N \equiv \inf\left\{t > 0 \middle| \left|X_t\right| \geq N\right\}, \tag{70}$$

and define the stopped process $X_t^N \equiv X_{t \wedge T_N}$. For $k \geq 3$,

$$\begin{aligned}\sum_{i=1}^{[nt]}\left|X_{i/n}^N - X_{(i-1)/n}^N\right|^k &= \sum_{i=1}^{[nt]}(2N)^k\left|\frac{X_{i/n}^N - X_{(i-1)/n}^N}{2N}\right|^k \\ &\leq (2N)^k\sum_{i=1}^{[nt]}\left|\frac{X_{i/n}^N - X_{(i-1)/n}^N}{2N}\right|^3 \\ &= (2N)^{k-3}\sum_{i=1}^{[nt]}\left|X_{i/n}^N - X_{(i-1)/n}^N\right|^3.\end{aligned} \tag{71}$$

Hence

$$\left|\sum_{k=1}^{M}\left(a_k\sum_{i=1}^{[tn]}\left(X_{i/n}^N - X_{(i-1)/n}^N\right)^k\right)\right| \leq \left\{\sum_{k=1}^{M}\left|a_k\right|(2N)^{k-3}\right\}\left\{\sum_{i=1}^{[tn]}\left|X_{i/n}^N - X_{(i-1)/n}^N\right|^3\right\}. \tag{72}$$

(68) implies that the first term on the right hand side tends to a finite limit as M tends to infinity. Denoting the limit by c_N,

$$\left|V^{(n)}\left(X^N, h\right)_t\right| \leq c_N\left\{\sum_{i=1}^{[nt]}\left|X_{i/n}^N - X_{(i-1)/n}^N\right|^3\right\} \text{ for all } N. \tag{73}$$

So for each N and for all $t < T_N$ the h-variation of X is bounded. Since X is continuous by assumption, its cubic variation converges to zero in probability as n tends to infinity (see Lepingle, 1976, and Jacod, 2008). Hence, by (73), the h-variation of X equals 0 on $[0, T_N]$, and the f-variation of X equals the quadratic variation on the same interval. Since T_N is an increasing sequence of stopping times tending to infinity, $T_N \wedge t$ increases to t a.s.. Furthermore, for $N' > N$, $V(X^N, f)_t = V(X^{N'}, f)_t$ on $\left[0, T_N \wedge t\right]$, hence the f-variation of X is well defined by the sequence $V(X^N, f)_t$ and equals the quadratic variation of X. The result holds true for any function f that satisfies (67) and (68). In particular, it holds for $f = L$ and $f = E$. ■

Proof of Proposition 7

$$
\begin{aligned}
&g^Q\left(s_T - s_0, v_T^E - v_0^E\right) = \sum_{t=0}^{T-1} g^Q\left(\delta s_{t+1}, \delta v_{t+1}^E\right) + Y \\
&\text{where } Y \equiv 6\sum_{u=1}^{T-1}\sum_{t=u}^{T-1}\left\{\frac{\delta E_u - \Delta_0 \delta S_u}{S_0}\frac{\delta S_{t+1}}{S_t} + \frac{\delta S_u}{S_0}\frac{\delta E_{t+1} - \Delta_t \delta S_{t+1}}{S_t}\right\}.
\end{aligned} \tag{74}
$$

is an algebraic identity. The first part of the proposition immediately follows by taking expectations at time 0, and applying the definitions of realized and true third moment. Let $x_{u+1} \equiv \dfrac{\delta E_{u+1} - \Delta_u \delta S_{u+1}}{S_u}$ and $r_{t+1} \equiv \delta S_{t+1}/S_t$. Approximate Y by:

$$
Y \approx 6\sum_{u=1}^{T-1}\sum_{t=u}^{T-1}\left\{\operatorname{cov}\left(x_u, r_{t+1}\right) + \operatorname{cov}\left(r_u, x_{t+1}\right)\right\} = 6\sum_{u=0}^{T-1}\sum_{t=0,\neq u}^{T-1}\operatorname{cov}\left(r_{t+1}, x_{u+1}\right). \tag{75}
$$

Now x_{u+1} is the return on a delta hedged option position with duration T-u-1

$$
\begin{aligned}
x_{u+1} &= \left(1 + r_{u+1}\right)\ln\left(1 + r_{u+1}\right) - r_{u+1} + \tfrac{1}{2}\left(1 + r_{u+1}\right)\delta v_{u+1}^E \\
&\approx \tfrac{1}{2}\left\{r_{u+1}^2 - \sigma_u^2 + \left(T - u - 1\right)\delta\sigma_{u+1}^2\right\}
\end{aligned} \tag{76}
$$

where σ_u is the daily implied Black-Scholes volatility of the Entropy Contract at time u. If shifts in the volatility term structure are parallel, then to remove the dependence of x_u on T define $y_t \equiv x_t / (T - t)$ and write

$$\begin{aligned} Y &\approx 3\sum_{u=0}^{T-1} \sum_{t=0,\neq u}^{T-1} (T-u-1)\operatorname{cov}(r_{t+1}, \delta y_{u+1}) \\ &= \tfrac{3}{2}\operatorname{var}(y)\operatorname{var}(r) \sum_{n=-T,\neq 0}^{T} (T-n-1)(T-|n|)\rho_{ry}(n) \text{ where } \rho_{r\sigma}(n) \equiv \rho(r_t, y_{t+n}) \end{aligned} \tag{77}$$

The realized third moment is simply the diagonal term where n=0, giving

$$\frac{TTM}{\mathbb{E}[RTM]} \approx \frac{\sum_{n=-T}^{T} (T-n-1)(T-|n|)\rho_{ry}(n)}{(T-1)T\rho_{ry}(0)} \tag{78}$$

The third part of the proposition comes from the identity

$$\begin{aligned} &g^V(s_T - s_0) = \sum_{t=0}^{T-1} g^V(\delta s_{t+1}) + Y \\ &\text{where } Y \equiv 2\sum_{u=1}^{T-1}\sum_{t=u}^{T-1} \frac{\delta S_u}{S_0}\frac{\delta S_{t+1}}{S_t}, \end{aligned} \tag{79}$$

and then reasoning is as before. ■

APPENDIX 2

Moments under the SVCJ model

Under the SVCJ model, the second moment of returns at horizon τ (where second moment is defined in the sense of Proposition 4) is a linear in the instantaneous variance v_t

$$
\begin{aligned}
&V^L(v_t) = (A+B)\tau + C(\tau); \quad \text{where} \\
&A = \theta,\, B = \lambda\left(\mu_V/\kappa + 2\left(\bar{\mu}_S - \mu_S\right)\right),\, C = \left(v_t - E[v]\right)\left(\frac{1-\exp(-\kappa\tau)}{\kappa}\right), \\
&\bar{\mu}_S = \exp\left(\mu_S + \sigma_S^2/2\right) - 1, \text{ and } E[v] = \theta + \lambda\mu_V/\kappa.
\end{aligned}
\tag{80}
$$

A is the expected diffusive variance rate, B is the expected jump variance rate. C reflects the dependence on the initial variance. The correction term C arises under the risk neutral measure because the conditional variance is a linear function of the initial instantaneous variance. Using the $\mathbb{P}$ measure parameters in (80) gives the true variance of returns, while using the $\mathbb{Q}$ parameters gives the implied variance. The expected realized variance is equal to the true variance. In computing the unconditional variance, v is replaced by its unconditional expectation under $\mathbb{P}$, so the C term drops out when computing the expected true variance or expected realized variance, but the term remains when computing the expected implied variance.

The unconditional third moment (defined in the sense of Proposition 6), takes the form

$$
\begin{aligned}
&T\bar{M} = D\tau + 3f(\tau)(G+H) \\
&\text{where} \quad f(\tau) = \left(\tau - \frac{1-\exp\left(-(\kappa-\rho\sigma_V)\tau\right)}{\kappa-\rho\sigma_V}\right)\Big/(\kappa-\rho\sigma_V); \\
&D = 6\lambda\left(\left(\mu_S + \sigma_S^2 - 2\right)\bar{\mu}_S + 2\mu_S + \sigma_S^2\right); \quad G = \rho\sigma_V\left(\theta + \lambda\mu_V/\kappa\right); \quad H = \lambda\mu_V\bar{\mu}_S.
\end{aligned}
\tag{81}
$$

The third moment is the sum of two terms: the skewness of jumps in the underlying (D) which is linear in horizon, and the covariation term which increases with f – initially quadratic in horizon and asymptotically linear for long horizons as mean reversion kicks in. f can be interpreted as the beta of the variance of returns over the horizon on the instantaneous (squared) volatility of the underlying; G and H capture the covariation between returns and instantaneous volatility, with G being the diffusive element and H being the jump component.

The true third moment is got by substituting the $\mathbb{P}$ parameters into equation (81). The realized third moment is got by using the $\mathbb{P}$ parameters everywhere except in f where the $\mathbb{Q}$ measure parameters are used. This reflects the fact that the realized third moment is computed using changes in implied variance rather than in the true variance. So the bias in the realized third moment comes entirely from the wedge in the value of f – the beta of variance on instantaneous volatility - under the two measures. In the SVCJ model, the beta depends on the speed of mean reversion κ. The higher the mean reversion, the lower the beta. With the speed of mean reversion being almost twice as high under the $\mathbb{Q}$ measure as under $\mathbb{P}$, this induces a substantial negative bias in the realized third moment, a bias that increases with the horizon.

References

Andersen, L. B. G. 2007. Efficient Simulation of the Heston Stochastic Volatility Model. Available at *SSRN*: http://ssrn.com/abstract=946405.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys. 2003. Modeling and Forecasting Realized Volatility. *Econometrica* 71:579-625.

Ang, A., J. Chen, and Y. Xing. 2006. Downside Risk. *Review of Financial Studies* 19: 1191-239.

Ang, A., R. Hodrick, Y. Xing, and X. Zhang. 2006. The Cross-Section of Volatility and Expected Returns. *Journal of Finance* 61:259–99.

Backus, D., M. Chernov, and I. Martin. 2011. Disasters Implied by Equity Index Options. *Journal of Finance* 66:1969-2012.

Bakshi, G., N. Kapadia, and D. Madan. 2003. Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options. *Review of Financial Studies* 16:101-43.

Bakshi, G., and D. Madan. 2000. Spanning and Derivative Security Valuation. *Journal of Financial Economics* 55:205-38.

Barro, R. 2009. Rare Disasters, Asset Prices, and Welfare Costs. *American Economic Review* 99:243-64.

Black, B. 1976. Studies of Stock Price Volatility Changes. *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics* 177–81.

Bondarenko, O. 2010. Variance Trading and the Market Price of Variance Risk. Available at SSRN: http://ssrn.com/abstract=1943254 or http://dx.doi.org/10.2139/ssrn.1943254

Britten-Jones, M., and A. Neuberger. 2000. Option Prices, Implied Price Processes, and Stochastic Volatility. *Journal of Finance* 55:839-66.

Broadie, M., M. Chernov, and M. Johannes. 2007. Model Specification and Risk Premia: Evidence from Futures Options. *Journal of Finance* 62:1453-90.

Brunnermeier, M., S. Nagel, and L. Pedersen. 2008. Carry trades and currency crashes. *National Bureau of Economic Research* Working Paper 14473.

Campbell., J.Y., A.W. Lo and A.C. MacKinlay. 1996. The Econometrics of Financial Markets. *Princeton University Press*.

Carr, P., and L. Wu. 2007. Stochastic Skew in Currency Options. *Journal of Financial Economics* 86:213-47.

Carr, P., and L. Wu. 2009. Variance Risk Premiums. *Review of Financial Studies* 22:1311-41.

Christie, A. C. 1982. The Stochastic Behavior of Common Stock Variances—Value, Leverage and Interest Rate Effects. *Journal of Financial Economics* 3:145–66.

Duffie, D., J. Pan, and K. Singleton. 2000. Transform Analysis and Option Pricing for Affine Jump-Diffusions. *Econometrica* 68:1343-76.

Harvey, C.R., and A. Siddique. 2000. Conditional Skewness in Asset Pricing Tests. *Journal of Finance* 55:1263-95.

Jacod, J. 2008. Asymptotic Properties of Power Variations and Associated Functionals of Semimartingales. *Stochastic Processes and their Applications* 118:517-59.

Jiang, G. J., and Y.S. Tian. 2005. The Model-Free Implied Volatility and Its Information Content. *Review of Financial Studies* 18:1305-42.

Kim, T.. and H. White. 2004. On More Robust Estimation of Skewness and Kurtosis. *Finance Research Letters* 1:56-73.

Kozhan, R., A. Neuberger, and P. Schneider. 2011. The Skew Risk Premium in Index Option Prices. *AFA 2011 Denver Meetings Paper*, available at http://ssrn.com/abstract=1571700.

Kraus, A., and R. H. Litzenberger. 1976. Skewness Preference and the Valuation of Risk Assets. *Journal of Finance* 31:1085-100.

Lepingle, D. 1976. La variation d'ordre p des semi-martingales. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete* 36:295-316.

Lo, A., and A. C. MacKinlay. 1988. Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test. *Review of Financial Studies* 1:41-66.

Longstaff, F., and M. Piazzesi. 2004. Corporate Earnings and the Equity Premium. *Journal of Financial Economics* 74:401-21.

Merton, R.C. 1969. Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case. *Review of Economics and Statistics* 51:247-57.

Rietz, T. 1988. The Equity Risk Premium a Solution. *Journal of Monetary Economics* 22:117-31.

Schoutens, W. 2005. Moment Swaps. *Quantitative Finance* 5:525-30.

Schoutens, W., E. Simons, and J. Tistaert. 2005. Model Risk for Exotic and Moment Derivatives. In Kyprianou, A.E. et al. (eds.) "*Exotic Option Pricing and Advanced Levy Models*", Wiley, 67-97.

Xing, Y., X. Zhang, and R. Zhao. 2010. What Does the Individual Option Volatility Smirk Tell Us About Future Equity Returns? *Journal of Financial and Quantitative Analysis* 45:641-62.

Table 1: Parameter estimates used in simulations

	γ	κ	θ	σ_V	μ_V	μ_S (%)	σ_S (%)	ρ	λ (%)
$\mathbb{P}$ measure	0.015	0.026	0.54	0.08	1.48	-2.63	2.89	-0.48	0.6
$\mathbb{Q}$ measure	0.031	0.057	0.246	0.08	8.78	-5.39	5.78	-0.48	0.6

The parameters for the SVCJ model $dS_t/S_t = \gamma dt + \sqrt{V_t} dW_t^S + \left(e^{Z_t^S} - 1\right) dN_t$ where V follows $dV_t = \kappa\left(\theta - V_t\right)dt + \sigma_V \sqrt{V_t} dW_t^V + Z_t^V dN_t$. Parameters are under the physical and risk-adjusted ($\mathbb{P}$ and $\mathbb{Q}$) measures. The numbers correspond to daily percentage returns and are taken from Tables I (row "EJP") and IV of Broadie, Chernov and Johannes (2007). The drift parameter γ is chosen to make S a martingale under both measures.

Table 2: Simulation Results

	True	Sample	Implied	Realized
		Monthly returns		
Second Moment	0.214	0.214	0.301	0.213
(x100)		(0.038)	(0.013)	(0.026)
Third Moment	-0.045	-0.044	-0.238	-0.037
(x1000)		(0.045)	(0.001)	(0.010)
Skewness	-0.454	-0.427	-1.594	-0.307
		(0.367)	(0.063)	(0.030)
		Annual Returns		
Second Moment	2.45	2.46	3.81	2.44
(x100)		(0.88)	(0.03)	(0.29)
Third Moment	-1.50	-1.49	-4.81	-0.81
(x1000)		(3.15)	(0.01)	(0.18)
Skewness	-0.391	-0.359	-0.648	-0.194
		(0.833)	(0.006)	(0.020)

The top panel reports the moments of monthly log returns generated by the stochastic volatility process with jumps described in equation (22). Moments are computed using the model free definitions from this paper. The first column contains the true moments of the physical distribution, and are obtained analytically. The next three columns show the moments computed from 10,000 simulations of the process. Each simulation comprises 200 months of 22 trading days. For each simulation, the second and third moments and skewness coefficient of the 200 monthly returns are computed. The average of these statistics across the 10,000 simulations are then recorded in the second column, together with the cross-sectional standard deviation across simulations in parentheses. The third and fourth columns do the same for the implied statistics (using one month options) and for the realized characteristics (computed using daily returns).

The lower panel is similar except that it shows statistics of annual returns computed over 20 year simulated histories. A year is defined as a period of 252 trading days.

Table 3: Realized and Implied Variance and Skewness of the S&P500 1997-2009

	Sample	Implied	Realized	Correl with 2^{nd} moment	
				Implied	Realized
			Monthly returns		
Second Moment (x100)	0.23	0.47 (0.53)	0.39 (0.63)		
Third Moment (x1000)	-0.16	-0.64 (1.36)	-0.38 (1.21)	-0.954	-0.934
Skewness	-1.98	-1.90 (0.88)	-1.10 (0.92)	0.297	0.074
			Three monthly returns		
Second Moment (x100)	0.82	1.44 (1.00)	1.16 (1.84)		
Third Moment (x1000)	-0.43	-2.99 (3.22)	-2.39 (7.53)	-0.952	-0.976
Skewness	-0.58	-1.69 (0.48)	-1.39 (0.50)	0.451	0.166
			Annual returns		
Second Moment (x100)	4.11	5.50 (3.17)	4.75 (4.76)		
Third Moment (x1000)	-6.21	-15.00 (15.13)	-18.14 (29.22)	-0.916	-0.954
Skewness	-0.74	-1.10 (0.35)	-1.60 (0.51)	0.235	0.403

The implied second and third moments are calculated from quoted option prices each month on the trading day following the previous option expiry date in December 1997 to September 2009 each month for monthly expiries, and three monthly for the longer maturities. The realized statistics are calculated using daily returns over the remaining life of the option. The sample statistics are computed using returns to expiry. Each cell shows the mean with the standard deviation in parentheses. Moments are computed using the model free definitions from this paper. The correlations are with the corresponding second moment – implied with implied, and realized with realized.

Table 4: Estimation of the Term Structure of the Realized Third Moment

n	*Log β_n*		*β_n*	*Relative Skew*
	Estimate	*Std error*		
3	0		1.00	1
6	0.777	0.054	2.17	1.122
9	1.030	0.063	2.80	1.150
12	1.163	0.074	3.20	1.147

The table reports the results of the panel regression $\log\left(Y_{t,n}/Y_{t,3}\right)=\sum_{i=6,9,12}\log\left(\beta_i\right)D_i+\tilde{\varepsilon}_{t+3,n}$

where D_i is a dummy that takes the value of 1 if $n = i$, and zero otherwise, $\tilde{\varepsilon}$ is an error term. $Y_{t,n}$ is the realized third moment (*RTM*) over the quarter starting at time *t*, computed using the implied variance of options that expire at time $t+n$, where time is measured in months. The standard errors are White cross-section standard errors. The relative skew is $\left(3/n\right)^{3/2}\sum_{j\le n}\beta_j$.

Table 5: Estimated Ratio of True Moment to Realized Moment

Horizon	Variance	Third Moment
1 month	0.71 [0.87,1.13]	0.77 [0.83,1.16]
3 months	0.67 [0.76,1.23]	0.88 [0.72,1.30]
1 year	0.88 [0.51,1.42]	1.01 [0.44,1.54]

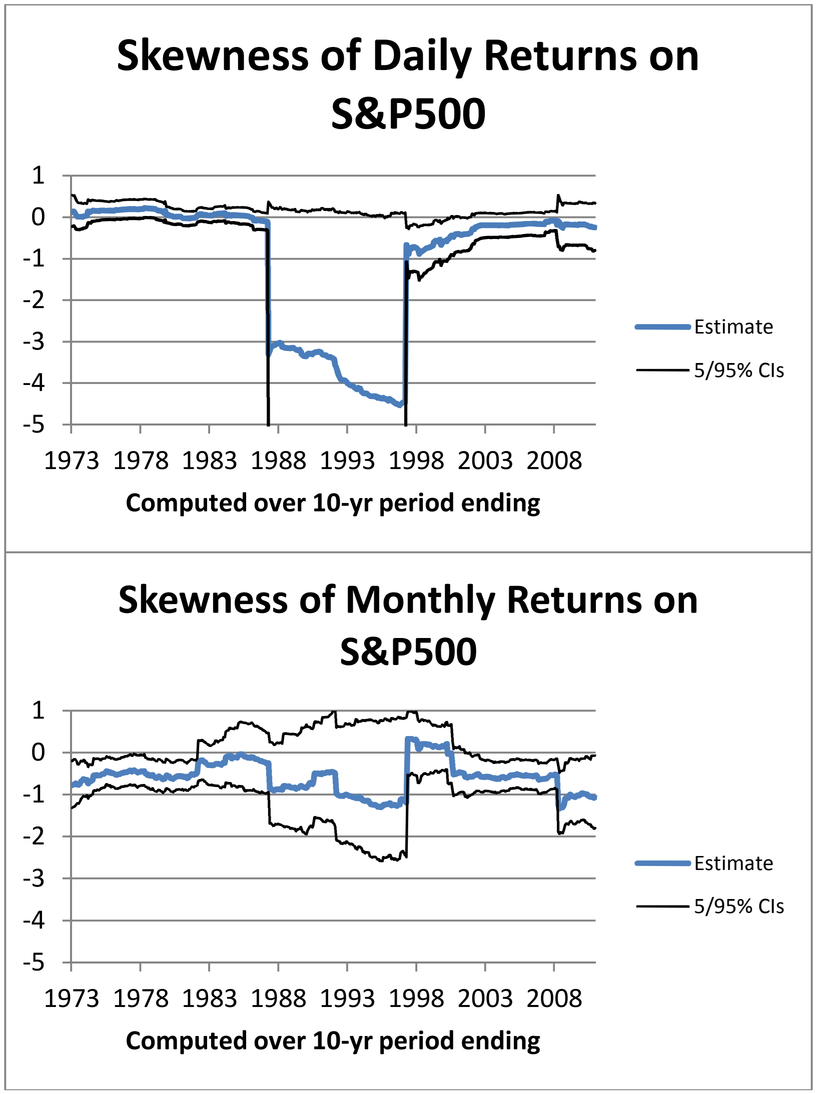
The estimates of the ratio of true variance and the third moment of returns to their realized counterparts make use of Proposition 7, and are computed from the auto- and cross-correlation structure of returns on the index and on hedged entropy contracts on the S&P500 over the period January 1996 to October 2010. The entropy contract return is based on a notional maturity equal to two thirds of the horizon, with prices obtained by linearly interpolating between the prices of the two nearest maturing entropy contracts. The numbers in square parentheses are the 5% and 95% intervals obtained by bootstrapping the returns data 1000 times. They show the confidence intervals under the null hypothesis of zero auto-correlation, and hence unbiased estimates.

Table 6: Regression of Realized Skew Coefficient on Implied

	α	β_1	β_2	**adjR^2**
1 month	-0.480	0.328		9.4%
	(-2.74)	(3.70)		
	-0.240	0.297	0.264	15.9%
	(-1.49)	(3.33)	(2.46)	
3 month	-0.595	0.470		19.1%
	(-2.61)	(3.16)		
	-0.541	0.359	0.161	19.0%
	(-2.70)	(2.55)	(1.48)	
6 month	-0.364	0.838		25.5%
	(-1.30)	(4.35)		
	-0.382	0.904	-0.084	24.2%
	(-1.34)	(3.65)	(-0.72)	
12 month	-0.857	0.672		19.4%
	(-2.87)	(3.01)		
	-0.658	0.657	0.152	20.2%
	(-2.41)	(3.21)	(0.98)	

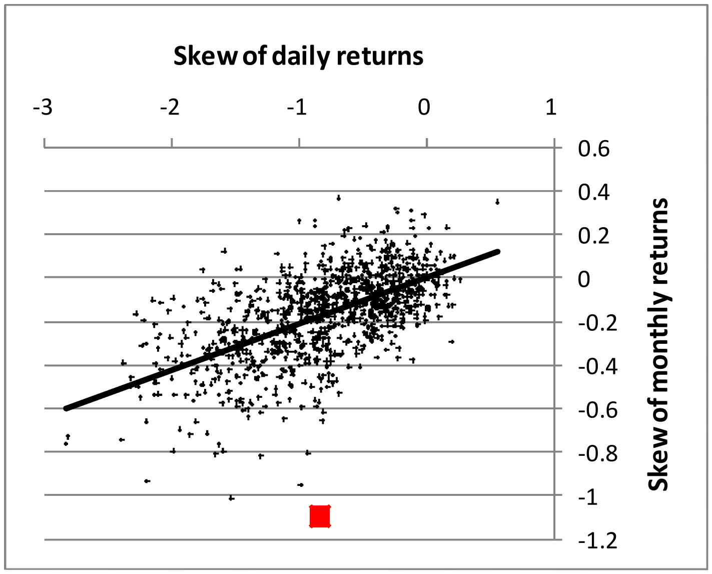
The table reports OLS results for the regression $RSC_{t,t+n} = \alpha + \beta_1 ISC_{t,t+n} + \beta_2 RSC_{t-m,t} + \varepsilon_{t+n}$ for n = 1, 3, 6 and 12 months, and m = min(n, 3) months. $RSC_{t,t+n}$ is the realized skew coefficient over the period (t, $t+n$) while $ISC_{t,t+n}$ is the implied skew coefficient computed using prices at time t on options that expire at time $t+n$. t–statistics in parentheses use Newey-West standard errors. Regressions are monthly for n =1, and quarterly otherwise, on the S&P500 index over the period 1998-2009.

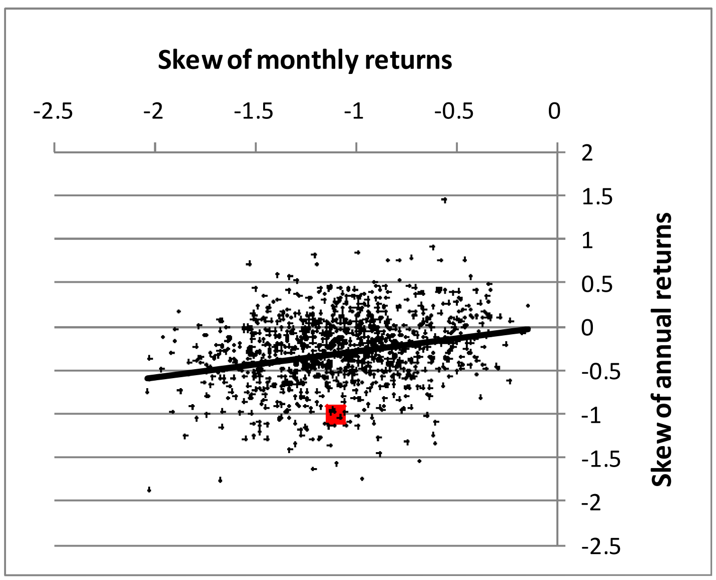
Figure 1: Skewness of S&P500 returns using overlapping 10 year periods



The charts show the skewness coefficient of the excess log returns on the S&P500 computed using rolling 10 year periods, starting in July 1963 and ending June 2011, together with the 5% and 95% confidence intervals. In the top panel, the skewness for each period is computed using daily returns over the period, while in the lower panel the computation is done using non-overlapping monthly returns. Confidence intervals are computed by bootstrapping 10,000 times.

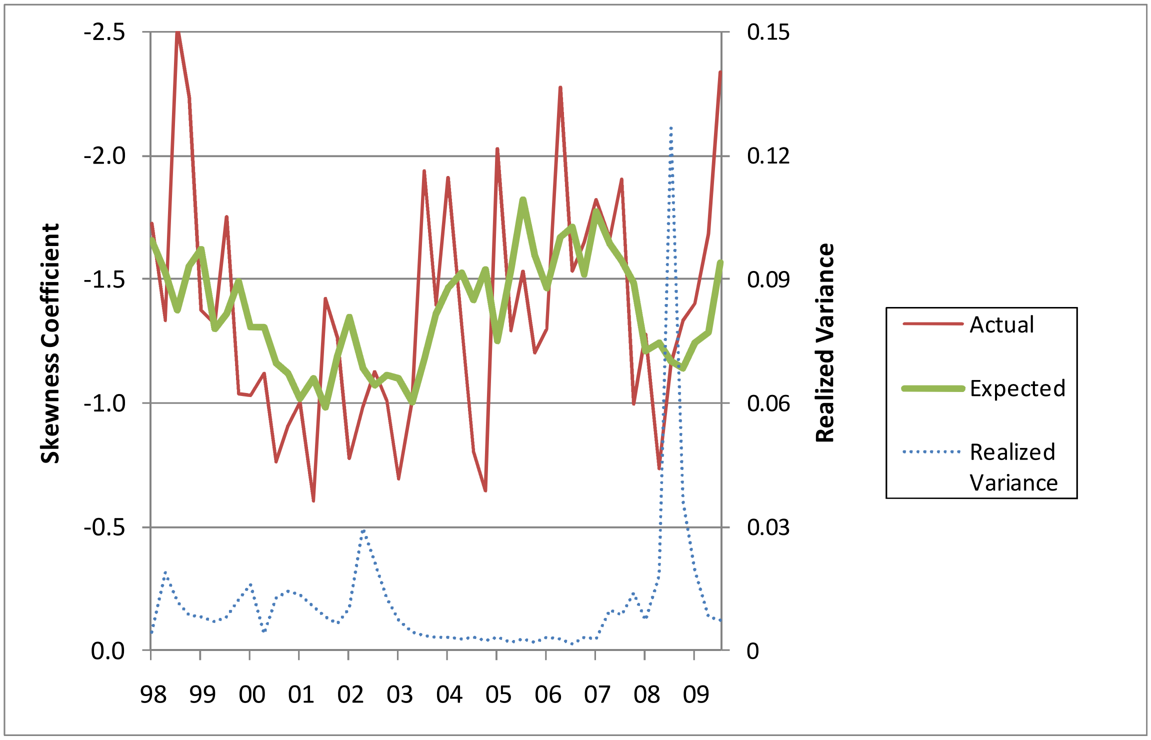
Figure 2:Skewness of index returns at different intervals





The upper plot uses bootstrapped daily returns to simulate the joint distribution of the skewness coefficient of daily and monthly log excess returns on the S&P500 index 1963-2011. The large square shows the actual skewness of daily and monthly returns over the period. The bold line plots the line $y = \sqrt{22}x$.The lower plot is similar except that it shows annual returns against monthly returns and the line is $y = \sqrt{12}x$.

Figure 3: Realized Skewness of 3 monthly returns on the S&P 500



The graph shows the realized skewness coefficient of 3 monthly returns on the S&P500 computed using daily returns and implied variance changes over successive quarters from March 1998 to December 2009, together with the expected skewness coefficient at the beginning of the quarter. The expected skew is computed using the model in Table 6, where expected realized skew is a linear function of lagged realized skew and implied skew. The realized variance over the quarter is plotted on the right hand axis for comparison purposes.

[1] I use the term "high frequency" to mean high frequency relative to the horizon rather than restricting it to intra-day returns.

[2] The data in this section are taken from the CRSP data base (June 2011). Excess log returns are defined as $\log(R^M/R^F)$ where R^M is the total daily return on the S&P500 and R^F is the total daily return on T-bills, computed from the monthly rate. p-values and the 95% confidence intervals reported below are obtained by bootstrapping.

[3] To simplify the algebra, the variance is not annualized; it tends to increase in magnitude with time to maturity.

[4] The key condition is that the predictable component of returns is predetermined. If volatility is stochastic then this condition is certainly violated under any pricing measure, and cannot be expected to hold under the physical measure, and the quadratic variation is not an unbiased estimator of the conditional variance of log returns. An implication of Propositions 4 and 5 together is that, if the price process is continuous, the quadratic variation of returns is an unbiased estimate of the logarithmic variance $\mathbb{E}_0\left[L\left(s_T - s_0\right)\right]$, under any pricing measure.

[5] I am much indebted to Dr Eberhard Mayerhofer for the proof of this Proposition.

[6] The QE algorithm recommended in Andersen (2007) is used for the discretization; the results are substantially the same if Euler discretization is used.

[7] In the entire sample there are eight out of 154 months in which the realized third moment is positive, and no cases at longer maturities.

[8] The figure of two thirds is chosen to ensure that the constant maturity contract has the same weighted average maturity as the entropy contracts used for estimating the third moment. The conclusions are not sensitive to the choice of maturity.

[9] The cubic is $f\left(t\right)=\frac{t}{2}\left\{-v^{L}+\left(2v^{L}-v^{E}\right)t+\left(v^{E}-v^{L}\right)t^{2}\right\}$.